

Matrix-Tensor Methods for Display Independent 3D Video Processing and Analysis

A Thesis

submitted by

SANTOSH KUMAR

*in partial fulfilment of the requirements
for the award of the degree of*

DUAL DEGREE

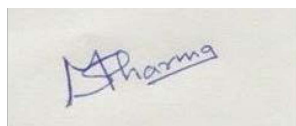


**DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.**

June 2020

PROJECT CERTIFICATE

This is to certify that the thesis entitled “**Matrix-Tensor Methods for Display Independent 3D Video Processing and Analysis**”, submitted by **Santosh Kumar** (EE15B110), to the Department of Electrical Engineering, Indian Institute of Technology, Madras, for the award of the **Dual Degree** (B.Tech. and M.Tech.), is a bonafide record of research work carried out by him under my supervision. The content of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any other degree or diploma.



27-6-2020

Dr. Mansi Sharma

Research Guide

INSPIRE Faculty

Dept. of Electrical Engineering

IIT Madras

Chennai - 600036, India

Place: Chennai

Date: 27th June 2020

ACKNOWLEDGEMENTS

During the course of my Dual Degree project, IIT Madras, I was very blessed to have constant and dynamic guidance from all my well-wishers. I would like to heartily thank each and every one of them for extending themselves for helping and encouraging me which has helped me to overcome all the obstacles in the path of progress of this project.

My first and foremost gratitude goes to **Dr. Mansi Sharma**, who guided me to carry on this work. She allowed me to be part of her group and work under her supervision. She is the source of steady encouragement, guidance and motivation.

My heartfelt gratitude to Balasubramanyam Appina who partly helped me in the successful completion of this work.

I would heartily extend my gratitude for Electrical Engineering Department, IIT Madras who have provided me all the required facilities for the completion of my project. My deep acknowledgement to all the faculty members of the Department of Electrical Engineering, IIT Madras for their support and guidance.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
LIST OF TABLES	iii
LIST OF FIGURES	1
1 Latent Factor Modeling of Users Subjective Perception for Stereoscopic 3D Video Recommendation	2
1.1 Introduction	2
1.2 Related Work	4
1.3 MATHEMATICAL MODELING OF S3D VIDEOS RECOMMENDATION SYSTEM	6
1.4 Results and Discussion	8
1.5 Conclusion	11
2 DEPTH VIDEO CODING	13
2.1 Introduction	13
2.2 HEVC: Brief Overview	15
2.3 Proposed Scheme	16
2.4 Quality Measurements	17
2.5 Conclusion	20
3 VIDEO PLUS DEPTH VIDEO CODING	21
3.1 Introduction	21
3.2 3D-HEVC: Brief Overview	22
3.3 Proposed Scheme	23
3.4 Quality Measurements	24
3.5 Conclusion	26

LIST OF TABLES

1.1	Performance evaluation of proposed algorithm on LFOVIAS3DPh2 and IRCCYN video dataset subjective scores.	10
2.1	PSNR vs Bitrate analysis of Camera4 Ballet and Breakdancers dataset	18
3.1	PSNR vs Bitrate analysis of Camera4 RGB Ballet and Breakdancers dataset	25
3.2	PSNR vs Bitrate analysis of Camera4 Depth Ballet and Breakdancers dataset	26

LIST OF FIGURES

1.1	Subjective score distribution of dataset. The μ , m and σ denote the statistical measures (mean (μ), median (m), standard deviation (σ)) of the subjective ratings.	6
1.2	Perceptual quality rates of subjective study and proposed objective study on pristine and distorted ‘Hall’ S3D videos.	9
1.3	Performance plot: a)RMSE error variation across epochs. b)Variation of proposed algorithm LCC score over 100 trails where Standard deviation of LCC score over 100 trails is 2×10^{-4}	11
2.1	Compact data representation	14
2.2	Typical HEVC video encoder (with decoder modeling elements shaded in light gray)	16
2.3	Overview of Encoder	16
2.4	Overview of Decoder	17
2.5	PSNR vs Bitrate plot for Ballet	19
2.6	PSNR vs Bitrate plot for Breakdancers	20
2.7	Visual quality comparison over synthesized view. (a) uncompressed depth image (b) depth images decoded by proposed HEVC (HM);	20
3.1	Example of typical 3D-HEVC encoding/decoding flow	23
3.2	Overview of Encoder	23
3.3	Overview of Decoder	23
3.4	PSNR vs Bitrate plot for Ballet RGB	27
3.5	PSNR vs Bitrate plot for Breakdancers RGB	28
3.6	PSNR vs Bitrate plot for Ballet Depth	28
3.7	PSNR vs Bitrate plot for Breakdancers Depth	29
3.8	Visual quality comparison over synthesized view. (a) uncompressed color image (b) color images decoded by proposed HEVC (HM);	29
3.9	Visual quality comparison over synthesized view. (a) uncompressed depth image (b) depth images decoded by proposed HEVC (HM);	29

CHAPTER 1

Latent Factor Modeling of Users Subjective Perception for Stereoscopic 3D Video Recommendation

ABSTRACT

Numerous stereoscopic 3D movies are released every year to theaters and created large revenues. Despite the improvement in stereo capturing and 3D video post-production technology, stereoscopic artifacts which causes viewer discomfort continue to appear even in high-budget films. Existing automatic 3D video quality measurement tools can detect distortions in stereoscopic images or videos, but they fail to consider viewer's subjective perception of those artifacts, and how these distortions affect their choices. In this paper, we introduce a novel recommendation system for stereoscopic 3D movies based on a latent factor model that meticulously analyse viewer's subjective ratings and influence of 3D video distortions on their preferences. To the best of our knowledge, this is a first-of-its-kind model that recommends 3D movies based on stereo-film-quality ratings accounting correlation between the viewer's visual discomfort and the stereoscopic-artifact perception. The proposed model is trained and tested on benchmark IRCCYN and LFOVIAS3DPh2 S3D video quality assessment datasets. The experiments revealed that resulting matrix-factorization based recommendation system is able to generalize considerably better for viewer's subjective ratings.

1.1 Introduction

The audience of 3D films and virtual reality content is growing, as most of the films or YouTube videos have been released in the stereoscopic 3D format today. There are three popular approaches to generate a stereoscopic 3D video (S3D): 1) Scene acquisition using a stereo camera, 2) 2D-to-3D video conversion, which means creation of left and right eye views from the original source video, 3) Rendering, which is the process of

synthesizing views by means of 3D reconstruction or employing global 3D models and computer vision techniques [1, 2, 3, 4].

Despite advancement in technology, there are numerous sources of visual artifacts to appear in the created stereoscopic picture/video [6, 7]. A comprehensive study of visual artifacts in S3D content has been carried out at MSU Graphics & Media Lab, Moscow State University, under VQMT3D project [6, 7] in cooperation with IITP RAS. The research study identified potential artifacts in several popular Hollywood S3D movies. The artifacts like disparity, scale, color, sharpness mismatches or temporal asynchrony, cardboard, crosstalk effects are prominent in the S3D 3DTV content. Besides, different types of artifacts at various stages of the content delivery affect S3D video. The compression, blur and frame-freeze distortions influence 3D video in format-conversion and representation stage, and in the coding and transmission stage [14]. Zeri and Livi [16] interviewed 854 people. They recognized frequent symptoms like eye strain, blurred vision and a burning sensation after watching 3D movies in theaters. Even high-budget films, like *Pirates of the Caribbean*, *Dolphin Tale*, *The Three Musketeers*, *The Avengers*, etc., contain scenes with geometric and color impairments, camera rotation difference, shift vertical variation between the left and right views. An important research study conducted by Miguel et al. [15] on 3D content using psychophysiological methods establish complex effects of visual discomfort over 3DTV viewer's emotional arousal, which leads to problems like headache, nausea, fatigue and eye strain, etc. The compression artifacts and their variation with a depth range on 3D displays noticeably affects viewer's perception [17, 18, 19].

The most reliable way to reduce such distortions is to correct and enhance the stereoscopic-content quality during production. But correction process is extremely labor intensive and heavily rely on degree of automation and on the workflow which is not cost efficient. The algorithms for automatic detection of such artifacts and quality assessment are emerging [20, 21]. However, measuring frequency and intensity of an artifact does not account how painful it can be for the viewer. Therefore, it is critical to consider subjective perception ratings for artifacts, that is, which egregious distortions affect a viewer notably and which distortions are within tolerable limits of his/her visual comfort.

In this paper, we proposed a novel recommendation system for S3D movies. The

well-controlled subjective experiments and careful statistical analysis conducted by most studies establish that discomfort is greater for some specific distortions than for others when viewing stereo video [5, 7]. Mainly the influence is from the content itself. We observed most significant information for designing a recommendation system for S3D movies is that describe the viewer's perceptual discomfort with the particular distortion types. Despite enough advances in image/video quality objective assessment techniques, it is difficult to propagate the same achievement for S3D video because automatic estimation of relevant characteristics for problems that causes visual discomfort is nontrivial. We wonder when even very simple yet reliable metrics measure several problems affecting stereo quality on the fly. Thus, it is crucial to account subjective ratings for healthy and reasonable 3D video watching as well as properly designing of recommendation system.

Our recommendation system is build on latent factor model, rely on viewer-movie ratings. Given a set of pristine and distorted S3D videos and their subjective ratings, our latent factor model that is based on matrix factorization map viewer's and 3D videos to a set of latent features. The problem of predicting perceptual quality rates of S3D video is formulated as a matrix completion problem for the user-movie rating matrix. Our system rate the S3D videos in accordance with the user's discomfort level. Our model recommendation mechanism can easily integrates within Netflix matrix factorization methods, which is most important class of collaborative filtering approaches. The proposed recommendation system will be very useful in reducing the flood of low-quality 3D content online by ratings stereo 3D more-consistent with quality. The encouraging results obtained by statistical analysis of the proposed model conducted on benchmark IRCCYN and LFOVIAS3DPh2 S3D data demonstrates its potential for generating accurate predictions.

1.2 Related Work

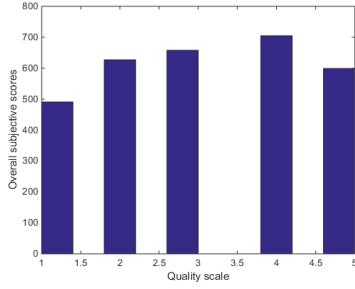
A comprehensive survey of algorithms used by Netflix for its Recommendation System is found in a paper written by Leidy Esperanza MOLINA FERNÁNDEZ [24]. It covered Collaborative Filtering, Content-based Filtering, model-based SVD, PCA, and Probabilistic Matrix Factorization techniques. The paper explains a movie recommen-

dition mechanism build within Netflix on the Matrix factorization (MF) approach that learns the latent preferences of users and movies from the ratings and make a prediction of the missing ratings using dot product of the latent factors [27, 24].

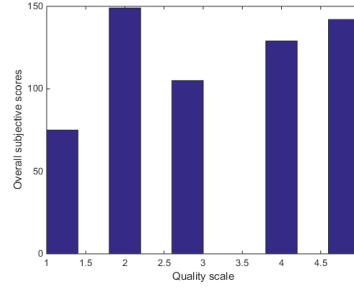
Lu et al. [25] applied MF model for computing vector representations of words. Their work demonstrated how a convolutional neural network can be integrated into MF model to produce interpretable recommendations. Lee et al. [26] demo model considered freshly uploaded YouTube videos. Here the collaborative filtering approach is not much applicable since it relies on aggregate user behavior. Instead, they modeled recommendation problem as a video content-based similarity learning. They learned deep video embeddings and predict ground-truth video relationships from trained model. However, this approach is built up purely based on video content signals.

YouTube provides a vast collection of 2D videos. In contrast since 2009, YouTube offers users interesting feature to upload two channel stereo videos for 3D viewing experience. YouTube flash players can support anaglyph videos in red/cyan, green/magenta or blue/yellow layout and follow row/column interlaced display on the screen. The 3D content on YouTube appear (or display) in accordance to the relevance order. Tsingalis et al. [22] presented a study on YouTube recommendation graphs of 2D and 3D videos. They studied the statistical relevance or recommendation properties of social network sites like Facebook, Tweeter and Flickr, such as power-law distribution. Also they analyzed clustering methods to understand the existence of media content groups. Davidson et al. [9] discussed in details about the recommendation system in use at YouTube. The study reveals YouTube recommends personalized sets of videos to users based on their previous activity on the web. They discussed some unique challenges YouTube faces for video endorsement and how to address them. Covington et al. [13] describe a YouTube system at a high level and center their study about substantial performance improvements brought by deep learning. They presented deep architecture built on deep candidate generation and separate ranking model for recommending YouTube videos.

Estrada and Simeone [8] developed a recommender system for guiding physical object substitution in virtual reality. This user-perception based recommender approach allows them to watch the physical world whilst navigating the virtual environment through a video feed. The user identifies the location of object placement in the surroundings given the feed.



(a) IRCCYN dataset.
 $\mu = 3.09$, $m = 3$, $\sigma = 1.35$.



(b) LFOVIAS3DPh2 dataset.
 $\mu = 3.19$, $m = 3$, $\sigma = 1.37$

Figure 1.1 Subjective score distribution of dataset. The μ , m and σ denote the statistical measures (mean (μ), median (m), standard deviation (σ)) of the subjective ratings.

Niu et al. [23] presented a video recommendation system based on the affective analysis of the users. Their subjective model evaluates feature of emotion fluctuation based on the Grey Relational Analysis (GRA). Certain video features are extracted and mapped to the well-known Lovheim emotion-space specifying prominent human feelings, patterns, attitudes and behaviour such as Anger, Distress, Surprise, Fear, Enjoyment, Shame, Interest, and Contempt. GRA-based recommendation method is developed under the Fisher model to analyse extracted emotions as factors.

Zhang et al. [11] developed a recommendation system for Mobile AR application incorporating user’s preferences, location and temporal information in an aggregated random walk algorithm. Their system predicts user’s preferences modifying the graph edge weight and computing the rank score. Similarly, Shi et al. [10] predicts individual location recommendation, Chatzopoulos and Hui [12] anticipates object recommendation in Mobile AR environments.

1.3 MATHEMATICAL MODELING OF S3D VIDEOS RECOMMENDATION SYSTEM

We proposed a novel recommendation system for stereoscopic 3D videos based on a Matrix Factorization (MF) model [27]. In the proposed model, viewer’s and S3D movies are mapped to a joint latent factor space. The row or column associated to a specific viewer or S3D movie is referred as the latent factors. In the mapped latent factor space of dimensionality, say f , the viewer-movie ratings are analyzed as inner products. Suppose each S3D movie i is associated with a latent vector $q_i^m \in R^f$, and each user

u is associated with a latent vector $p_j^u \in R^f$. In the proposed problem formulation, for a given movie i , the elements of q_i^m estimate the extent to which the S3D movie holds those factors, whether distorted with a particular artifact or free from that. For a given user u , the elements of p_j^u determine the extent of user acceptance has in S3D movies that are high on the corresponding factors, again, whether distorted with a particular artifact or not. The model approximates viewer u 's rating of S3D movie i by measuring resulting dot product, $\hat{r}_{ui} = q_i^{m^T} p_j^u$. The dot product captures interconnection between the viewer u and S3D movie i , that is, the viewer's overall acceptance/tolerance in the particular distortion affecting the movies. Once the mapping is computed for each S3D movie and viewer to factor vectors $q_i^m, p_j^u \in R^f$, the proposed model easily determines the rating a viewer will give to any S3D movie with distortions by using \hat{r}_{ui} .

We avoided imputation in proposed model [28]. The observed ratings are modeled directly as suggested by [27, 29] and avoided overfitting through the regularization. On the set of known matrix ratings, regularized squared error is minimized to learn the factor vectors q_i^m, p_j^u as

$$\min_{\hat{p}, \hat{q}} \sum_{(u,i) \in \mathfrak{S}} (r_{ui} - q_i^{m^T} p_j^u)^2 + \lambda(\|q_i^m\|^2 + \|p_j^u\|^2) \quad (1.1)$$

where, \mathfrak{S} is the training set of (u, i) pairs for which r_{ui} is known.

To make matrix factorization approach more effective in our proposed application-specific requirements, we add biases in capturing the full ratings of the observed signals

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^{m^T} p_j^u \quad (1.2)$$

The observed rating in (1.2) is broken down into its four components: global average (or mean), 3D movie bias, viewer bias, and viewer-movie interaction. This allows each component to represent only the part of an observed signal relevant to it. The model is learned by minimizing the squared error function as

$$\min_{\hat{p}, \hat{q}, \hat{b}} \sum_{(u,i) \in \mathfrak{S}} (r_{ui} - \mu - b_i - b_u - q_i^{m^T} p_j^u)^2 + \lambda(\|q_i^m\|^2 + \|p_j^u\|^2 + b_i^2 + b_u^2) \quad (1.3)$$

The stochastic gradient descent algorithm [27, 30, 31] is used to optimize equation (1.3). For better accuracy in prediction, the algorithm loops through all ratings in the

training data and estimate the model parameters. The system estimates \hat{r}_{ui} for each given training case. The prediction error is determined as

$$E_{ui} = r_{ui} - \mu - b_i - b_u - q_i^{mT} p_j^u \quad (1.4)$$

The parameters are updated as

$$b_i \leftarrow b_i + \varsigma(E_{ui} - \lambda b_i) \quad (1.5)$$

$$b_u \leftarrow b_u + \varsigma(E_{ui} - \lambda b_u) \quad (1.6)$$

$$q_i^m \leftarrow q_i^m + \rho(E_{ui} p_j^u - \lambda q_i^m) \quad (1.7)$$

$$p_j^u \leftarrow p_j^u + \rho(E_{ui} q_i^m - \lambda p_j^u) \quad (1.8)$$

where, ρ and ς specify constant magnitudes that accounts proportion by which parameters are modified in the opposite direction of the gradient.

The objective of our matrix factorization model is to predict the unknown future S3D video ratings, from the learned model obtained by fitting the earlier observed ratings. We determined the regularization constant λ by cross-validation [32].

1.4 Results and Discussion

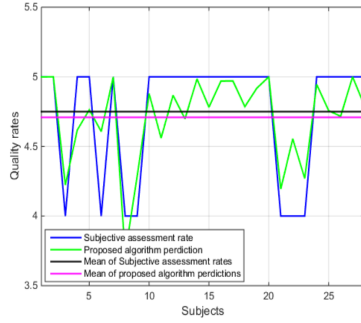
The efficacy of the proposed algorithm is evaluated on the IRCCYN [33] and LFOVIAS3DPh2 [20] S3D video datasets. IRCCYN database has 10 reference and 100 test S3D video sequences. The video sequences have a resolution of 1920×1080 and saved in .avi container. The frame rate is 25 fps and a duration of either 16 sec or 13 sec. The database is a combination of H.264 and JP2K, scaling and down sampling distorted S3D video sequences. These artifacts are applied symmetrically on each view of an S3D video and published the DMOS scores as subjective scores. Human assessment on perceptual quality was performed in single stimulus continuous quality evaluation (SSCQE) with hidden reference method. They have used 5 scales to rate the perceptual quality of an S3D video and 28 subjects involved in the study. They have published each subject quality score and an overall mean quality score of the dataset. LFOVIAS3DPh2 S3D video dataset has 12 pristine sequences with good variety of structure, texture, depth



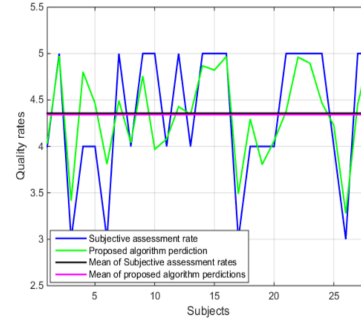
(a) Pristine stereoscopic video frame.



(b) Distorted stereoscopic video frame.



(c) Pristine stereoscopic video.



(d) Distorted stereoscopic video.

Figure 1.2 Perceptual quality rates of subjective study and proposed objective study on pristine and distorted ‘Hall’ S3D videos.

and temporal information. The video sequences have a resolution of 1920×1080 and duration of 10 seconds with a frame speed of 25 fps. They created 288 test stimuli by introducing the H.264 and H.265 compression, Blur and Frame freeze distortions. The dataset is a combination of symmetric and asymmetric S3D videos. They have used SSCQE method to perform the subjective study and 20 subjects involved in the study. They published each subject perceptual quality score and final difference mean opinion score of the dataset. Figure 1.1 shows the subjective score distribution of IRCCYN and LFOVIAS3DPh2 S3D video datasets. From the plots, it is clear that both the datasets are diverse in video perceptual quality range. Also, it is clear that the subjective ratings are consistent and followed the trend observed in perceptual quality of S3D videos.

Figure 1.2a shows the 1st frame from the left view of the ‘Hall’ S3D video from IRCCYN dataset. Figure 1.2b shows the 1st frame of H.264 (quantization parameter = 38) compressed S3D video of the corresponding reference view. Figures 1.2c and 1.2d show the distribution of subjective assessment rates and proposed algorithm predicted perceptual quality rates of a pristine and distorted S3D videos, respectively. From the plot it is clear that the proposed algorithm accurately predicts the subjective quality rates of pristine and distorted videos. Also, the deviation between average scores of subjective rates and the proposed algorithm predictions is very less. The plot clearly

demonstrates the proposed algorithm efficacy to model the perceptual subjective quality ratings of a given video.

The performance of the proposed algorithm is measured using the Linear Correlation Coefficient (LCC), Spearman Rank Order Correlation Coefficient (SROCC) and Root Mean Square Error (RMSE). LCC indicates the linear dependence between two quantities. The SROCC measures monotonic relationship between two input sets. RMSE measures the magnitude error between estimated scores and subjective scores. Higher LCC and SROCC values indicate good agreement between subjective and objective measures, and lower RMSE signifies more accurate prediction performance. For both the databases, 80% of the human opinion scores are used for proposed algorithm training and the remaining samples are used for testing. In other words, the training and test sets are obtained by partitioning the set of available human opinion scores in the 80:20 proportion. We performed the random assignment for 100 trials of each epoch for statistical consistency, and calculated the mean of the LCC, SROCC and RMSE measures of each epoch to report the performance analysis. Table 1.1 shows the performance evaluation of the proposed algorithm on the training and test sets of IRCCYN and LFOVIAS3DPh2 S3D video datasets. It is clear that the proposed algorithm shows robust performance across all datasets.

Figure 1.3b shows the LCC score variation of 100 iterations of an epoch. From the plot it is clear that the scores are consistent across all iterations, and further, we experienced the lower standard deviation (2×10^4) of 100 LCC scores. Figure 1.3a shows the average training and test RMSE measure variation over 200 epochs. From the plot, it is clear that both the RMSE errors reduced with the number of epochs. These plots clearly demonstrate the proposed algorithm efficacy to estimate the human assessment quality of a given video.

Table 1.1 Performance evaluation of proposed algorithm on LFOVIAS3DPh2 and IRCCYN video dataset subjective scores.

Score	Training Set			Testing Set		
	LCC	SROCC	RMSE	LCC	SROCC	RMSE
IRCCYN	0.8873	0.8858	0.6903	0.8753	0.8700	0.7527
LFOVIAS3DPh2	0.8966	0.8911	0.5359	0.7385	0.7288	0.7849

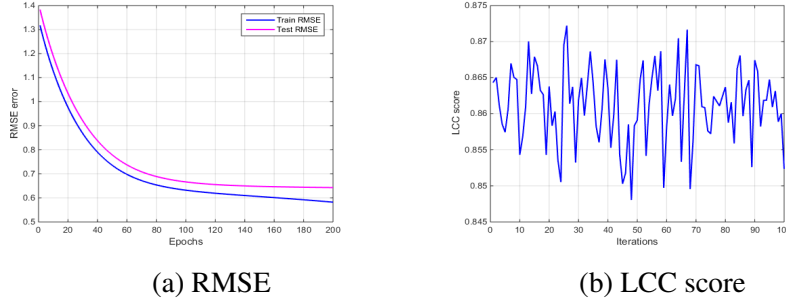


Figure 1.3 Performance plot: a)RMSE error variation across epochs. b)Variation of proposed algorithm LCC score over 100 trails where Standard deviation of LCC score over 100 trails is 2×10^{-4} .

1.5 Conclusion

This paper presented a novel recommendation system for S3D movies. This is a first attempt that accounts 3DTV viewer’s subjective ratings for visual artifacts and analyse their degree of visual discomfort to predict “rating” or “preference” that the viewer’s would give to the S3D movie. In this study, we considered four common distortion types; Blur, Frame-freeze, H.264 and H.265 compression; that adversely affect S3D video signal at different stages of the content generation and delivery chain. Experimental results on 3DTV viewer’s subjective study and parameter evaluation of latent factors demonstrate that the proposed matrix factorization based model improve accuracy of S3D video affective analysis and performance of recommendation. This model will be very useful for media-service providers like Netflix, Amazon, TiVo to recommend quality 3D videos and minimize flood of low-quality content based on the viewer’s subjective perception, depending on their age groups and preferences.

We will further extend this recommendation system by considering the detail analysis of commercial S3D movies. The model will be improved by offering per-frame analysis of artifacts causing potential visual discomfort while viewing stereo films like large horizontal disparity, vertical parallax, crosstalk noticeability, cardboard effect, stuck-to-background objects, stereo window violation, depth continuity, etc. Such artifacts earn poor rating according to the existing metrics. Combining objective and subjective scores will help reduce the error rate further while recommending new stereo movies.

Besides, we will perform affective analysis on the emotional reactions of 3DTV viewers while watching stereo 3D movies or virtual reality S3D content. We will account both subjective scores and brain-activity measurements to understand the depen-

dencies between the degree of viewer discomfort and the intensity of the distortions. This will help to better classify viewers from different age groups by their susceptibility to artifacts and movies content types. How this affect viewer's accumulation of discomfort caused by stereoscopic movies and influence recommendation ratings is an interesting endeavour of future study ?. To continue to work on this idea, we will account the percentage of viewers susceptible to various distortions. We will design new experiments and work on evaluation models like probabilistic matrix factorization (PMF) to improve the predictive accuracy. We will experiments on the linear combination of predictions of multiple PMF models with predictions of Restricted Boltzmann Machine (RBM) models. This could significantly improve the accuracy of the blended solution.

CHAPTER 2

DEPTH VIDEO CODING

ABSTRACT

The growing interest and popularity in 3D television had created new needs of storing the information related with 3D views. 3D video coding includes the use of multiple color views and depth maps associated to each view. An adequate coding of depth maps should be adapted which allows us to store 3D video efficiently. This paper proposes efficient techniques to compress a depth video. To increase efficiency first depth video is approximated by Tensor based method of Tucker decomposition with the Alternating Least Squares (ALS) algorithm and then efficiently compressed using HEVC encoder. Experimental results show the proposed scheme outperforms HEVC software on benchmark “Ballet” and “Breakdancers” sequences and Kinect depth sequences. Experimental results showed the effectiveness of our method and it can be applied to interactive video coders

2.1 Introduction

Advances in computer graphics, computer vision, multimedia and related technologies together with increased interest in three-dimensional (3D) video technologies have promoted the development of new means to store and transmit video information. Depth images/video that indicates the real world distances enables lots of applications. It has shed light on the advancement of many applications in computer vision, such as robot navigation, gesture recognition, 3d reconstruction, human-machine interaction, pose tracking, activity detection, foreground/background segmentation, and so on. In 3D video areas, besides multi-view video, the typical representations of 3D video include video-plus-depth, multi-view video plus depth (MVD) and layered depth video (LDV), where the depth information permits the easier rendering of new views [38][39]. Recently, many depth cameras, such as Kinect [40] prevail and make the access of depth

image much easier. In Kinect, depth image is obtained based on disparity estimation over the pattern emitted from one infrared camera and that actually observed by another camera. Another type of depth camera, called Time-of-Fight camera, sense the depth by computing the time of flight from light being emitted to that light being received.

3D video, however, involves a huge amount of data that needs to be encoded and transmitted. Consequently, it is essential to have efficient 3D content representation and compression techniques in order to enable prospective 3D services and technologies. Depth maps are used to render new images and not to be viewed directly by the user. Thus, the aim when coding depth maps is to maximize the perceived visual quality of the rendered virtual color views instead of the visual characteristics of decoded depth maps themselves. Conventional image or video compression techniques have been designed for high visual quality, and are not well adapted to depth video coding.

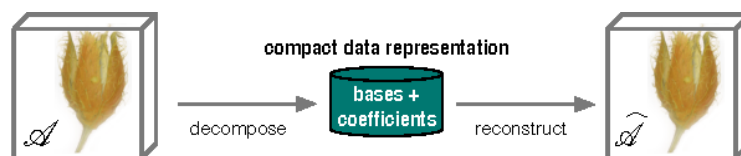


Figure 2.1 Compact data representation

Data approximation is widely used in the fields of computer graphics and scientific visualizations. One way to achieve data approximation is to decompose the data into a more compact and compressed representation. The general idea of a compact data representation is to express a dataset by a set of bases, which are used to reconstruct the dataset to its approximation when needed(see fig. 1). Precisely speaking, a set of bases usually consists of the actual bases and coefficients describing the relationship between the original data and the actual bases. Typically, such bases sets constitute less data than the original dataset, capture the most significant features, and, moreover, describe the data in a format that is convenient/appropriate for adaptive data loading.

Bases for compact data representation can be classified into two different types: pre-defined and learned bases. Predefined bases comprise a given function or filter, which is applied to the dataset without any a priori knowledge of the correlations in the dataset. In contrast, learned bases are generated from the dataset itself. Established

examples of pre-defined bases are the Fourier transform (FT) and the Wavelet transform (WT). Well-known examples of learned bases are the PCA or the SVD. Using pre-defined bases is often computationally cheaper, while using learned bases requires more computing time (to generate the bases), but potentially removes more redundancy from a dataset. Generally, PCA-like methods are able to extract the main data direction of the dataset and represent the data in a different coordinate system such that it makes it easier for the user to find the major contributions within the dataset. To exploit this, PCAs higher-order extension – tensor approximation (TA) – can be used for multidimensional datasets. A very popular numerical method to compute the decomposition for a given tensor Tucker decomposition and it should be observed that ALS is commonly-used for the Tucker decomposition and seems to be efficient for compression problems.

2.2 HEVC: Brief Overview

HEVC signifies a number of advances in video coding development. In the family of video coding standards, HEVC has the promise and potential to replace/supplement all the existing standards (MPEG and H.26x series including H.264/Advanced Video Coding AVC). Its video coding layer design is based on conventional block-based motion compensated hybrid video coding concepts. The HEVC standard is built to achieve multiple goals, including coding efficiency, ease of transport system integration and data loss resilience, as well as implementability using parallel processing architectures. In HEVC, the main goal was to achieve a compression gain higher when compared to the H.264/AVC at the same video quality. While the complexity of the HEVC encoder is several times that of H.264/AVC, the decoder complexity is within the range of the latter, making HEVC decoding in software very practical on current hardware[41]. HEVC retains the basic hybrid architecture of prior video coding standards such as H.264/AVC[42]

Source video, consisting of a sequence of video frames, is encoded or compressed by a video encoder to create a compressed bit stream, which will be stored or transmitted. Compared to AVC, HEVC provides the following new features [43] : quad-tree partitioning for prediction and transform with more and larger block sizes parallel process-

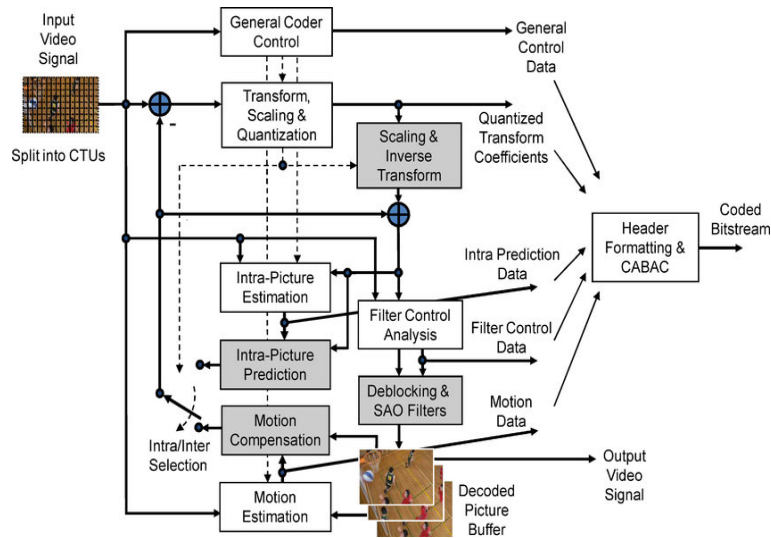


Figure 2.2 Typical HEVC video encoder (with decoder modeling elements shaded in light gray)

ing with tiles and wave-forms, ultra-low delay processing with dependent slices, inter-picture prediction block merging, advanced motion vector prediction, high throughput transform coefficient coding, transform skip mode for screen content coding, sample adaptive offset in loop filtering[44]

The improved coding efficiency of HEVC, however, come with a price tag: increased computational complexity. Compared with its predecessor, HEVC is complex for encoding and decoding. The HEVC standard is a general one suitable for the compression of all kinds of video.

2.3 Proposed Scheme



Figure 2.3 Overview of Encoder

Considering the special characteristics of depth videos, we propose a novel compression framework, aiming to enhance the coding efficiency while preserving the inherent depth features. Compression standards such as H.264/AVC and HEVC provide

superior coding performance through the exploration of the spatial and temporal correlations. For depth images/video, they can efficiently remove the redundancy in data.



Figure 2.4 Overview of Decoder

Fig. 3 & 4 shows the architecture of the proposed encoder and decoder. By using the framework of the conventional HEVC, it is efficient to implement the proposed depth encoder and decoder by utilizing Tensor ALS.

The encoder contains a pre-processing block that enables the spatial resolution and dynamic range reduction of depth signal, if necessary, for an efficient depth map compression. The motivation is that with an efficient Tensor algorithm, encoding the approximated depth data on can reduce the bit rate substantially while still achieving a good synthesized view quality. For the decoding process the framework remains as conventional HEVC decoder.

2.4 Quality Measurements

We take our experiments at the benchmark of HEVC test model (HM) using software of HM2.0 . We have conducted a series of experiments to evaluate the performance of the proposed depth compression techniques. We have tested with the Breakdancers and Ballet test sequences with resolutions of 1024×768 , of which both the color video and depth map are provided from Microsoft Research [45].

The bitrate of the compressed depth videos and the peak signal-to-noise ratio (PSNR) of the rendered virtual views are the two main performance measures for comparison.

Table 2.1 PSNR vs Bitrate analysis of Camera4 Ballet and Breakdancers dataset

	QP	BALLET		BREAKDANCERS	
		BITRATE (kbps)	PSNR (dB)	BITRATE (kbps)	PSNR (dB)
RANK 1	2	88.484	71.5652	78.7052	71.5823
	6	45.868	63.7623	40.6212	64.0371
	10	23.9336	60.5351	22.6816	60.962
	14	14.0212	58.3281	12.59	58.758
	20	7.5296	55.5065	6.5572	56.0513
	26	4.332	52.8418	3.6	53.3778
	38	2.0824	45.727	1.566	46.1376
RANK 5	2	989.8124	61.8702	772.6688	63.1238
	6	438.0408	58.4929	363.9332	59.7206
	10	218.046	56.504	186.9116	57.6155
	14	116.0116	54.8954	98.7312	55.8834
	20	50.9192	52.7956	40.8132	53.7284
	26	23.5004	50.1449	17.922	51.2406
	38	6.208	42.9849	4.9632	44.0549
RANK 10	2	1076.96	61.1878	1007.117	61.6846
	6	524.2048	58.1217	498.8484	58.5806
	10	282.5764	56.1354	268.8484	56.531
	14	162.8956	54.5264	151.2096	54.8959
	20	79.9164	52.3321	69.6912	52.81
	26	39.17	49.4471	32.9508	50.1857
	38	10.7896	41.9825	8.7676	42.6037
RANK 15	2	1198.818	60.6656	1119.452	61.0293
	6	591.5128	57.6406	561.506	58.0322
	10	328.4748	55.7001	311.1148	56.0421
	14	197.5608	54.1194	181.318	54.4764
	20	100.874	51.8871	88.4972	52.3299
	26	50.7916	48.8881	43.0976	49.5284
	38	13.9664	41.2267	11.5904	41.9059
RANK 20	2	1317.327	60.2517	1217.69	60.5687
	6	651.2208	57.2305	616.1936	57.6338
	10	367.1012	55.3158	347.8264	55.683
	14	222.6292	53.753	208.5464	54.1591
	20	114.3328	51.4555	105.37	51.9982
	26	57.136	48.3824	52.4428	49.099
	38	15.0484	40.7415	14.1104	41.3569

The empirical results, against different values of quantization parameter (QP) for different Tensor Rank, regarding the bitrate of the compressed depth video and the PSNR of the rendered virtual view are listed in Table 2.1.

Fig 2.5 and 2.6 shows the rate-distortion(RD) curves for ranks 1,5,10,15,20 in terms of the depth bitrate(dB) and the depth quality for Ballet and Breakdancers respectively.

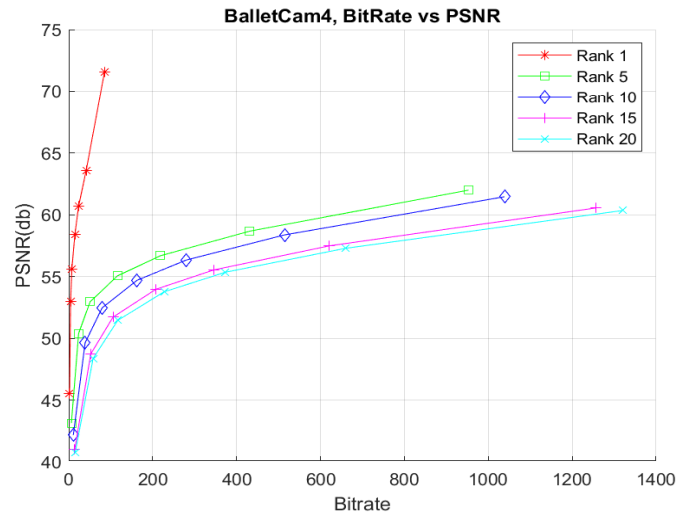


Figure 2.5 PSNR vs Bitrate plot for Ballet

Figs. 2.7 (a) and (b) show sample frames of the views generated based upon the reconstructed depths shown in Fig. 2.7 (a) (i.e. without the proposed encoder) and Fig. 2.7 (b) (i.e. with the proposed encoder), respectively.

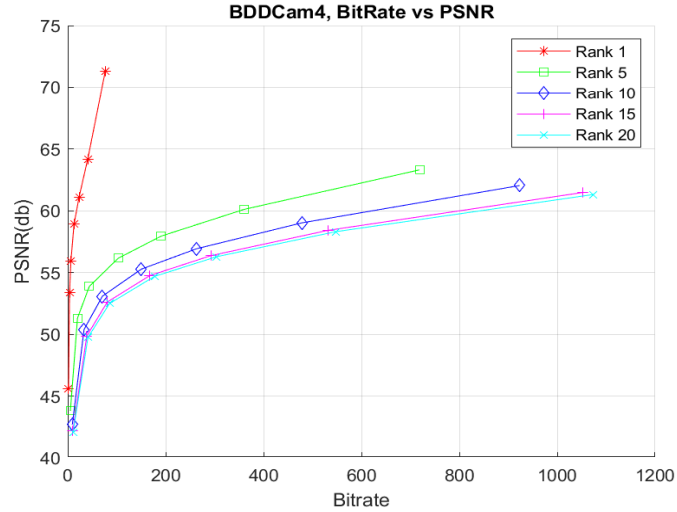


Figure 2.6 PSNR vs Bitrate plot for Breakdancers

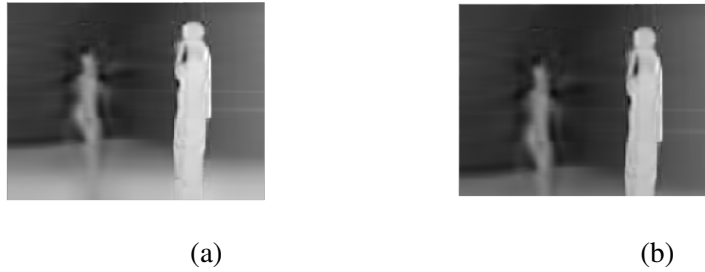


Figure 2.7 Visual quality comparison over synthesized view. (a) uncompressed depth image (b) depth images decoded by proposed HEVC (HM);

2.5 Conclusion

We have presented novel techniques to compress the depth video by using HEVC Framework and utilizing Tensor ALS to approximate the input to HEVC. The experimental results have shown the performance of the proposed scheme for different rank and different quantisation parameter (QP). It has been observed that as rank increases the bitrate of the compressed depth videos increases and the peak signal to-noise ratio (PSNR) of the rendered virtual views decreases. Also as QP increases both the bitrate of the compressed depth videos and the peak signal to-noise ratio (PSNR) of the rendered virtual views decreases. As a result, incurring lower coding bit rate, we can achieve the same quality of the synthesized view.

CHAPTER 3

VIDEO PLUS DEPTH VIDEO CODING

ABSTRACT

We propose in this paper a new method of Multi view videos coding based on the 3D-HEVC framework to simultaneously compress a color video and corresponding depth map. A popular format for 3D uses a conventional color video and an associated per sample depth map. 3D-HEVC has received a remarkable response due to its high compression efficiency which is based on High Efficiency Video Coding (HEVC). However, the complexity of its encoding process is large. We propose an efficient method in which First depth video and color video is approximated by Tensor based method of Tucker decomposition with the Alternating Least Squares (ALS) algorithm and then efficiently compressed simultaneously using 3D-HEVC encoder. The simulation results on Ballet and Breakdancers datasets show that the proposed method achieves better saving in depth bit-rate and depth PSNR compared with conventional 3D-HEVC based coding of MVD representations.

3.1 Introduction

Several high-resolution video applications have arisen in the last decade demanding high efficiency and quality of encoding. Besides, these videos are stored in several media and places and streamed over several heterogeneous communication systems distributed at the internet. Therefore, video coding experts spent a high effort in the standardization of the modern Two-Dimensional (2D) video coding standards such as High Efficiency Video Coding (HEVC)[46], VP9 [47], Audio Video Coding Standard 2 (AVS2) [48], to obtain a high encoded video quality with a reduced stream size. However, currently, video coding utilization goes beyond capturing and encoding simple 2D scenes. Now, video applications allow sharing screens or enjoying a three-dimensional (3D) experience that goes beyond 2D videos by providing a depth perception of the

scene. The 2D video coding standards do not encode these new video properties properly because they focus on the texture aspects of the scene; consequently, reducing the efficiency on capturing depth aspects of each video's scene. To fulfill this requirement, several HEVC extensions were designed by the video coding experts, including HEVC Screen Content (HEVC-SCC) [49], which enables achieving a higher performance when sharing computer screens or similar videos, and the 3D High Efficiency Video Coding (3D-HEVC) [50, 51], which better encode 3D video redundancies.

3.2 3D-HEVC: Brief Overview

In 2010, MPEG and VCEG established a Joint Collaborative Team on Video Coding (JCT-VC) to develop the HEVC standard and the first standard of HEVC was finalized on January 25, 2013. 3D-HEVC is a codec for 3D video which represents an extension of HEVC. It was put forward by JCT-3V in 2012 and hasn't been finalized yet. After 12 meeting discussions since 2012, there are more than 10 drafts and 15 test models for 3D-HEVC. The 3D-HEVC codec scheme came into being with many additional depth map coding tools and multi-view coding methods integrated into the HEVC codec.

The highest advantage of MVD is the stream size reduction for encoding a 3D video because the decoder can synthesize high-quality intermediary views interpolating texture views based on the depth data and using Depth Image Based Rendering (DIBR) [10] or other view synthesis techniques. These techniques allow synthesizing several high-quality intermediary texture views of the scene, reducing the number of stored/transmitted views. Figure 3.1 illustrates an abstraction of 3D-HEVC usage composed by coding and decoding processes. The first step is the scene capturing, then, the obtained data passes by the 3D-HEVC video encoding, followed by the video storing/transmitting. Next, the 3D, stereo and 2D video decoders decode the bit-stream of the encoded video. Finally, according to the output video format, the intermediary virtual views are synthesized.

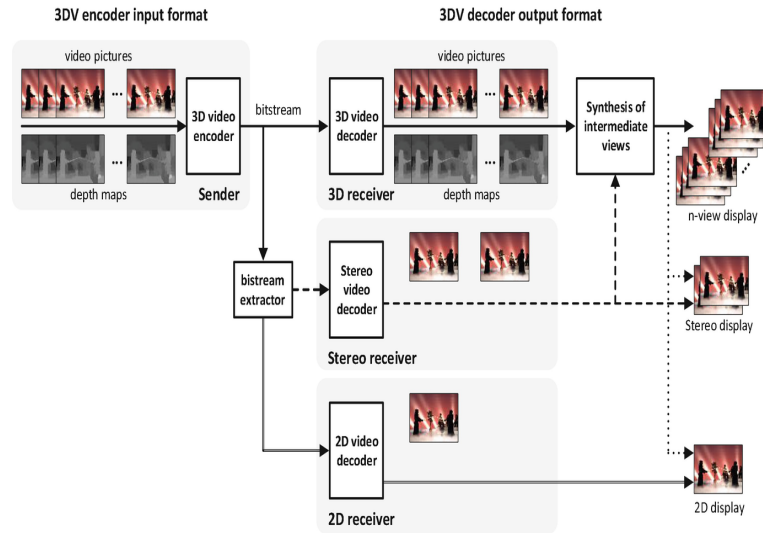


Figure 3.1 Example of typical 3D-HEVC encoding/decoding flow

3.3 Proposed Scheme

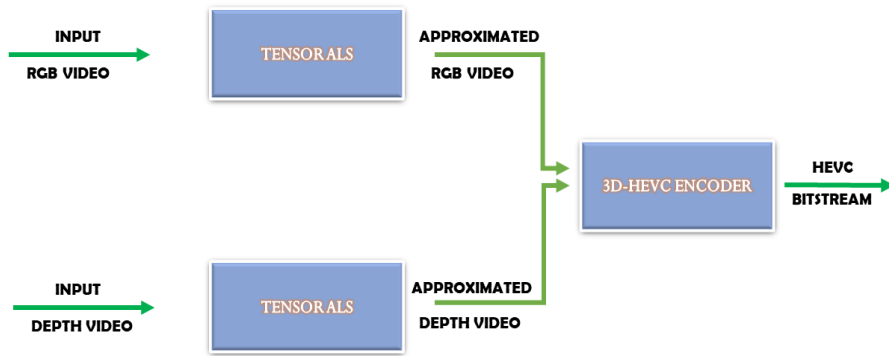


Figure 3.2 Overview of Encoder

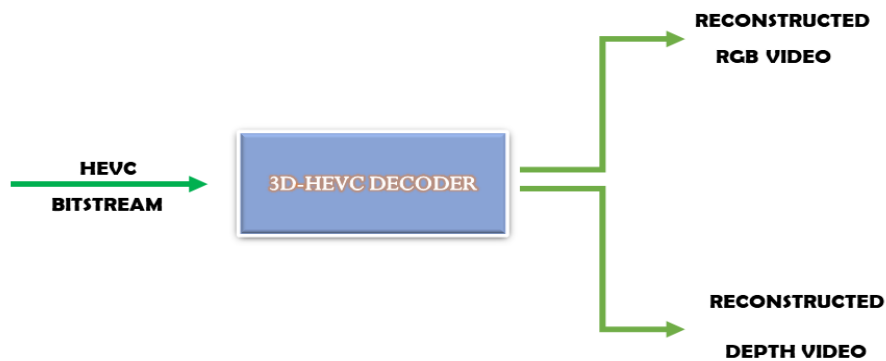


Figure 3.3 Overview of Decoder

Fig.3.2 & 3.3 shows the architecture of the proposed encoder and decoder. By using the framework of the conventional 3D-HEVC, it is efficient to implement the proposed encoder and decoder by utilizing Tensor ALS. The encoder contains a pre-processing block that enables the spatial resolution and dynamic range reduction of

depth signal, if necessary, for an efficient depth map compression. The motivation is that with an efficient Tensor algorithm, encoding on the approximated color video and corresponding depth data can reduce the bit rate substantially while still achieving a good synthesized view quality. For the decoding process the framework remains as conventional 3D-HEVC decoder.

3.4 Quality Measurements

Similar to previous chapter, We have conducted a series of experiments to evaluate the performance of the proposed color and depth compression techniques. We have tested with the Breakdancers and Ballet test sequences with resolutions of 1024×768 , of which both the color video and depth map are provided from Microsoft Research.

The bitrate of the compressed depth videos and the peak signal-to-noise ratio (PSNR) of the rendered virtual views are the two main performance measures for comparison. The empirical results, against different values of quantization parameter (QP), regarding the bitrate of the compressed color and depth video and the PSNR of the rendered views are listed in Table 3.1 and 3.2 respectively.

Fig 3.4 and 3.5 shows the rate-distortion(RD) curves for ranks 1,5,10,15,20 in terms of the color bitrate(dB) and the color quality for Ballet and Breakdancers respectively.

Fig 3.6 and 3.7 shows the rate-distortion(RD) curves for ranks 1,5,10,15,20 in terms of the depth bitrate(dB) and the depth quality for Ballet and Breakdancers respectively.

Figs. 3.8 show sample frames of the views generated based upon the reconstructed color frames shown in Fig. 3.8(a) (i.e. without the proposed encoder) and Fig. 3.8(b) (i.e. with the proposed encoder), respectively.

Figs. 3.9 show sample frames of the views generated based upon the reconstructed depths shown in Fig. 3.9(a) (i.e. without the proposed encoder) and Fig. 3.9(b) (i.e.

Table 3.1 PSNR vs Bitrate analysis of Camera4 RGB Ballet and Breakdancers dataset

	QP	BALLET		BREAKDANCERS	
		BITRATE (kbps)	PSNR (dB)	BITRATE (kbps)	PSNR (dB)
RANK 1	2	2500.445	62.3403	2266.011	62.7013
	6	796.4528	56.6491	811.6128	57.3797
	10	161.4368	54.6474	136.8992	55.1352
	14	72.2464	53.5462	61.8528	54.1431
	20	29.7264	51.7731	26.392	52.6401
	26	14.6048	49.1946	12.1952	50.3563
	38	5.2864	42.8713	4.9296	44.0849
RANK 5	2	3646.282	60.3261	4034.835	60.55
	6	1293.554	54.7863	1466.725	54.9954
	10	424.3536	52.8804	455.8688	53.1772
	14	225.84	51.5869	246.008	52.1379
	20	106.2528	49.4523	115.0848	50.4619
	26	50.0352	46.401	53.3024	47.838
	38	13.0112	39.44	13.6864	41.154
RANK 10	2	4101.798	59.5655	4430.072	59.5885
	6	1578.19	54.2794	1733.525	54.4752
	10	607.192	52.3914	659.6112	52.7316
	14	345.1472	51.0189	389.4864	51.6085
	20	167.48	48.7172	195.176	49.6426
	26	77.9648	45.4927	91.304	46.7025
	38	18.6608	38.3859	21.984	39.4506
RANK 15	2	4374.002	59.3158	4637.085	59.1582
	6	1756.597	54.0383	1921.637	54.2468
	10	732.0032	52.1496	819.7904	52.5061
	14	422.8144	50.7343	495.9344	51.264
	20	205.8384	48.3087	253.5392	49.0906
	26	94.1296	44.9975	118.8416	45.8254
	38	21.856	37.9686	26.9984	38.372
RANK 20	2	4561.434	59.1801	4960.086	58.8385
	6	1896.28	53.9328	2122.394	53.9992
	10	831.8608	52.0007	956.0816	52.2789
	14	488.4784	50.5145	594.5984	51.0259
	20	236.2176	47.9777	307.3104	48.7588
	26	106.392	44.587	143.712	45.3429
	38	23.6768	37.5252	32.2608	37.8289

with the proposed encoder), respectively.

Table 3.2 PSNR vs Bitrate analysis of Camera4 Depth Ballet and Breakdancers dataset

	BALLET			BREAKDANCERS	
	QP	BITRATE (kbps)	PSNR (dB)	BITRATE (kbps)	PSNR (dB)
RANK 1	2	166.8064	69.0769	157.5744	69.3469
	6	44.4	58.012	45.4336	58.306
	10	29.128	56.2816	26.5616	56.4007
	14	20.5664	54.7613	18.4944	55.0043
	20	12.6784	52.2568	11.1552	52.7243
	26	8.3488	49.3881	7.4864	49.7688
	38	4.5344	39.9835	3.936	41.2293
RANK 5	2	918.1616	57.2403	786.5744	58.6763
	6	347.0048	54.8219	295.376	55.8724
	10	182.5632	53.3963	155.5888	54.4284
	14	107.9296	52.1252	89.6032	53.2151
	20	55.536	49.7844	44.3136	51.1098
	26	29.3696	46.6961	23.832	48.0949
	38	8.8416	38.1427	7.6768	39.7496
RANK 10	2	1060.662	56.9473	1004.717	57.7897
	6	459.7184	54.5548	421.3712	55.1089
	10	260.9792	53.0213	241.0944	53.6417
	14	164.2368	51.7028	149.6432	52.3961
	20	87.4016	49.2054	78.2112	50.0284
	26	46.496	45.8725	41.3808	46.8986
	38	13.4096	37.5046	12.1504	38.4379
RANK 15	2	1277.701	56.1778	1142.682	57.154
	6	585.2512	53.8696	483.944	54.6057
	10	353.4224	52.3543	283.4368	53.1792
	14	228.9424	50.9753	179.5584	51.9398
	20	124.144	48.2813	94.3056	49.4873
	26	65.2656	44.7386	49.36	46.1873
	38	17.84	36.3202	13.6192	37.7684
RANK 20	2	1363.334	55.9977	1187.954	57.0947
	6	651.4288	53.7205	514.0944	54.5471
	10	399.8784	52.1495	305.9408	53.1235
	14	262.112	50.7106	195.0336	51.8574
	20	140.2656	47.8633	102.088	49.3243
	26	72.0768	44.2568	52.7904	45.9527
	38	19.0112	35.9311	14.3488	37.6808

3.5 Conclusion

We have presented novel techniques to compress the depth video by using 3D-HEVC Framework and utilizing Tensor ALS to approximate the input video sequences. The experimental results have shown the performance of the proposed scheme for different rank and different quantisation parameter (QP). It has been observed that as rank in-

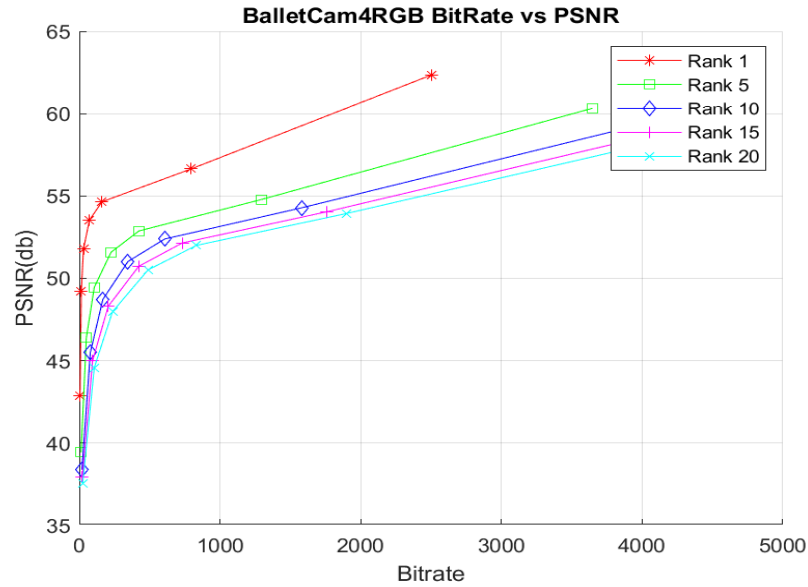


Figure 3.4 PSNR vs Bitrate plot for Ballet RGB

increases the bitrate of the compressed RGB and depth videos increases and the peak signal to-noise ratio (PSNR) of the rendered virtual views decreases. Also as QP increases both the bitrate and the peak signal to-noise ratio (PSNR) of the rendered virtual views decreases. As a result, incurring lower coding bit rate, we can achieve the same quality of the synthesized view. Experimental results showed the effectiveness of our method and it can be applied to interactive video coders

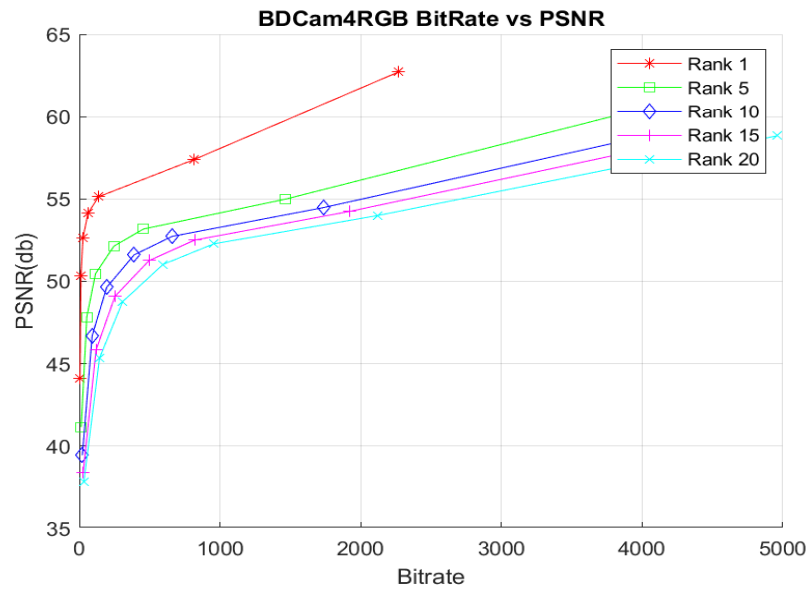


Figure 3.5 PSNR vs Bitrate plot for Breakdancers RGB

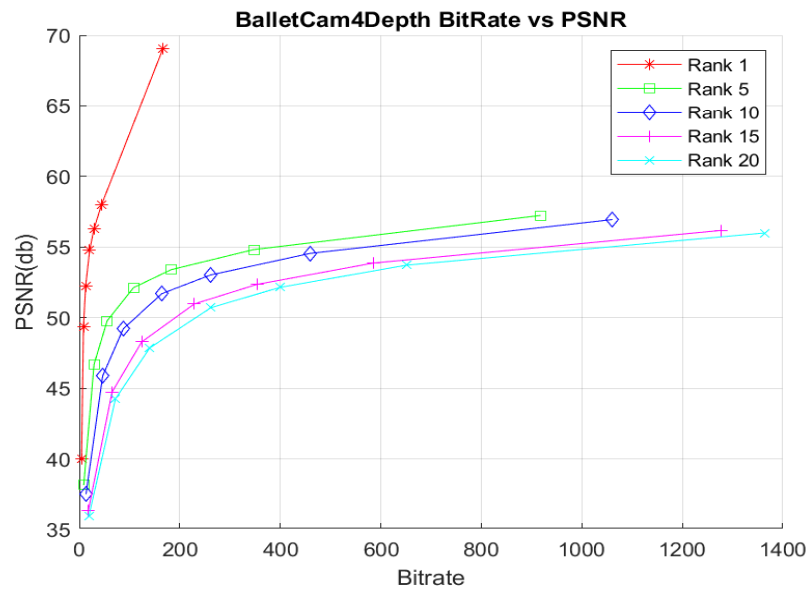


Figure 3.6 PSNR vs Bitrate plot for Ballet Depth

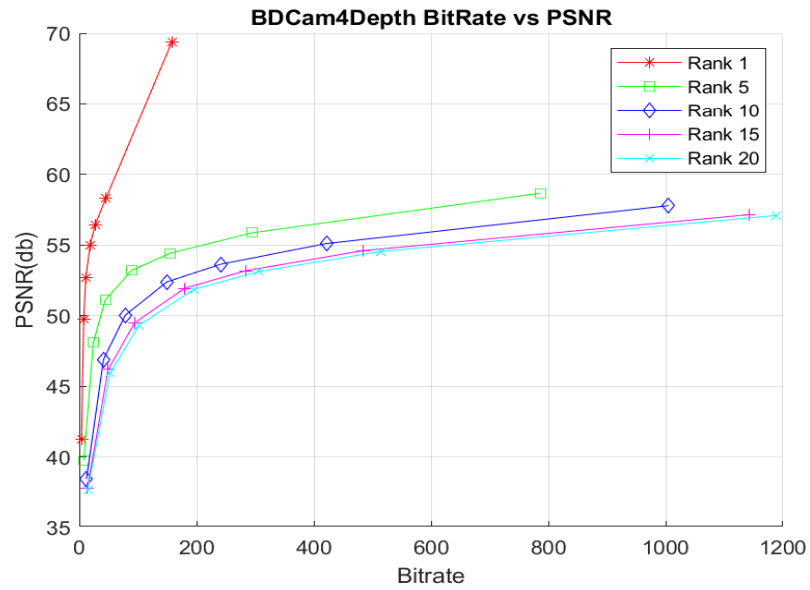


Figure 3.7 PSNR vs Bitrate plot for Breakdancers Depth

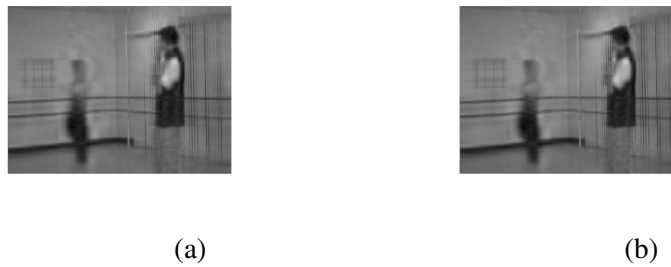


Figure 3.8 Visual quality comparison over synthesized view. (a) uncompressed color image (b) color images decoded by proposed HEVC (HM);



Figure 3.9 Visual quality comparison over synthesized view. (a) uncompressed depth image (b) depth images decoded by proposed HEVC (HM);

REFERENCES

- [1] A. Smolic et al. Three-Dimensional Video Postproduction and Processing. *In Proceedings of the IEEE*, vol. 99, no. 4, pp. 607-625, April 2011.
- [2] M. Sharma, S. Chaudhury and B. Lall. A Novel Hybrid Kinect-Variety-Based High-Quality Multiview Rendering Scheme for Glass-Free 3D Displays. *in IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2098-2117, Oct. 2017.
- [3] M. Sharma, Uncalibrated Camera Based Content Generation for 3D Multi-view Display, *Phd Dissertation*, Indian Institute of Technology Delhi, 2017.
- [4] Mansi Sharma, Santanu Chaudhury, Brejesh Lall, and M. S. Venkatesh. 2014. A flexible architecture for multi-view 3DTV based on uncalibrated cameras. *Journal of Visual Communication and Image Representation*, 25, 4 (May 2014), 599-621.
- [5] Jing Li, Marcus Barkowsky, Patrick Le Callet. Visual Discomfort in 3DTV: Definitions, Causes, Measurement, and Modeling. *Novel 3D Media Technologies* pp. 185-209, 2014.
- [6] A. Antsiferova and D. Vatolin. The influence of 3D video artifacts on discomfort of 302 viewers. *International Conference on 3D Immersion (IC3D)*, Brussels, 2017, pp. 1-8.
- [7] MSU 3D-video Quality Analysis, Video Quality Measurement Tool 3D Project, MSU Graphics & Media Lab (Video Group), <http://compression.ru/video/vqmt3d/>
- [8] J. Garcia Estrada and A. L. Simeone. Recommender system for physical object substitution in VR. *IEEE Virtual Reality (VR)*, Los Angeles, CA, 2017, pp. 359-360.
- [9] Davidson et al. The YouTube video recommendation system. *In Proceedings of*

- the fourth ACM conference on Recommender systems (RecSys '10)*, ACM, New York, NY, USA, 293-296.
- [10] Z. Shi, H. Wang, W. Wei, X. Zheng, M. Zhao and J. Zhao. A Novel Individual Location Recommendation System Based on Mobile Augmented Reality. *International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI)*, Beijing, 2015, pp. 215-218.
- [11] Zhuo Zhang, Shang Shang, Sanjeev R. Kulkarni, and Pan Hui. Improving augmented reality using recommender systems. *In Proceedings of the 7th ACM conference on Recommender systems (RecSys'13)*, ACM, New York, NY, USA, 173-176.
- [12] Dimitris Chatzopoulos and Pan Hui. ReadMe: A Real-Time Recommendation System for Mobile Augmented Reality Ecosystems. *In Proceedings of the 24th ACM international conference on Multimedia (MM'16)*, ACM, New York, NY, USA, 312-316.
- [13] Paul Covington, Jay Adams, and Emre Sargin. Deep Neural Networks for YouTube Recommendations. *In Proceedings of the 10th ACM Conference on Recommender Systems (RecSys'16)*. ACM, New York, NY, USA, 191-198.
- [14] A. Gotchev, G. B. Akar, T. Gapin, D. Strohmeier, A. Boev. Three-Dimensional Media for Mobile Devices. *Proceedings of IEEE*, Vol. 99, No. 4, pp. 708-741, April 2011.
- [15] M. Barreda-Ángeles, R. Pépion, E. Bosc, P. Le Callet and A. Pereda-Baños. How visual discomfort affects 3DTV viewers' emotional arousal. *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Budapest, 2014, pp. 1-4.
- [16] Fabrizio Zeri and Stefano Livi. Visual discomfort while watching stereoscopic three-dimensional movies at the cinema. *Ophthalmic and Physiological Optics*, 35(3):271-282, 2015.
- [17] G. Sanchez, J. Silveira, L. Agostini and C. Marcon. Performance Analysis of Depth Intra Coding in 3D-HEVC. *in IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

- [18] S. Jumisko-Pyykkö, T. Haustola, A. Boev, A. Gotchev. Subjective Evaluation of Mobile 3D Content: Depth Range versus Compression Artefacts. *Proceedings of SPIE, Multimedia for Mobile Devices, part of Electronic Imaging Symposium*, 2011.
- [19] B. Appina, K. Manasa, and S. S. Channappayya. Subjective and objective study of the relation between 3D and 2D views based on depth and bitrate. *In Electronic Imaging*, vol. 2017, no. 5, pp. 145-150, 2017.
- [20] A. Bokov, D. Vatolin, A. Zachesov, A. Belous, and M. Erofeev. Automatic detection of artifacts in converted S3D video. *In Proc. SPIE 9011, Stereoscopic Displays and Applications XXV*, vol. 9011, pp. 901112-1–901112-14, March 2014.
- [21] B. Appina, S. V. R. Dendi, K. Manasa, S. S. Channappayya and A. C. Bovik. Study of Subjective Quality and Objective Blind Quality Prediction of Stereoscopic Videos. *in IEEE Transactions on Image Processing*, 2019.
- [22] I. Tsingalis, I. Pipilis and I. Pitas. A statistical and clustering study on Youtube 2D and 3D video recommendation graph. *In 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Athens, 2014, pp. 294-297.
- [23] J. Niu, S. Wang, Y. Su and S. Guo. Temporal Factor-Aware Video Affective Analysis and Recommendation for Cyber-Based Social Media. *In IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 3, pp. 412-424, 2017.
- [24] Leidy Esperanza MOLINA FERNÁNDEZ. Recommendation System for Netflix. *Faculty of Science Business Analytics*, Vrije Universiteit Amsterdam, 2018.
- [25] Yichao Lu, Ruihai Dong, and Barry Smyth. Convolutional Matrix Factorization for Recommendation Explanation. *In Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, ACM, New York, NY, USA, Article 34, 2 pages, 2018.
- [26] Joonseok Lee, Nisarg Kothari, Paul Natsev. Content-based Related Video Recommendations. *Advances in Neural Information Processing Systems (NIPS) Demonstration Track*, 2016.

- [27] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix Factorization Techniques for Recommender Systems. *Computer*, Volume 42, Issue 8, August 2009, pp. 30-37
- [28] Dimitris Bertsimas, Colin Pawlowski, and Ying Daisy Zhuo. From predictive methods to missing data imputation: an optimization approach. *Journal of Machine Learning Research*, Res. 18, 1 (January 2017), 7133-7171.
- [29] A. Paterek. Improving Regularized Singular Value Decomposition for Collaborative Filtering. *In Proc. KDD Cup and Workshop*, ACM Press, 2007, pp. 39-42.
- [30] Bangti Jin and Xiliang Lu, On the regularizing property of stochastic gradient descent, *Inverse Problems*, Volume 35, Number 1, 2018.
- [31] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. *In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'08)*. ACM, New York, NY, USA, 426-434.
- [32] Ruslan Salakhutdinov and Andriy Mnih. Probabilistic Matrix Factorization. *In Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07)*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis (Eds.). Curran Associates Inc., USA, 1257-1264.
- [33] M. Urvoy et al. NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences. *Fourth International Workshop on Quality of Multimedia Experience*, Yarra Valley, VIC, 2012, pp. 109-114.
- [34] Y. Zhang, H. Zhang, M. Yu, S. Kwong and Y. Ho Sparse Representation-Based Video Quality Assessment for Synthesized 3D Videos. *in IEEE Transactions on Image Processing*, vol. 29, pp. 509-524, 2020, doi: 10.1109/TIP.2019.2929433.
- [35] Dumić, E., Sakic, K. and da Silva Cruz, L.A. Crowdsourced subjective 3D video quality assessment *Multimedia Systems*, 25, 673–694 (2019).
- [36] Fan, Q., Luo, W., Xia, Y. et al. metrics and methods of video quality assessment: a brief review. *Multimed Tools Appl*, 78,31019–31033 (2019).

- [37] Peng, Zongju and Wang, Shipei and Chen, Fen and Zou, Wenhui and Jiang, Gangyi and Yu, Mei. (2019). Quality Assessment of Stereoscopic Video in Free Viewpoint Video System. *Journal of Visual Communication and Image Representation*, 63, 10.1016/j.jvcir.2019.06.011.
- [38] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz. Coding algorithms for 3dtv—a survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11):1606–1621, 2007.
- [39] Philipp Merkle, Aljosa Smolic, Karsten Muller, and Thomas Wiegand. Multi-view video plus depth representation and coding. volume 1, pages I – 201, 09 2007.
- [40] Wikipedia contributors. Kinect — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Kinect&oldid=958970348>, 2020. [Online; accessed 1-June-2020].
- [41] F. Pescador, M. Chavarrias, M. J. Garrido, E. Juarez, and C. Sanz. Complexity analysis of an hevc decoder based on a digital signal processor. *IEEE Transactions on Consumer Electronics*, 59(2):391–399, 2013.
- [42] Telecom ITU. Advanced video coding for generic audiovisual services. *ITU-T Recommendation H.264*, 2003.
- [43] Dragorad Milovanovic and Z. Bojkovic. Mpeg video deployment in interactive multimedia systems: Hecv vs. avc codec performance study. *WSEAS Transactions on Signal Processing*, 9:167–176, 01 2013.
- [44] Z. M. Miličević and Z. S. Bojković. Hecv performance analysis for hd and full hd applications. In *2014 22nd Telecommunications Forum Telfor (TELFOR)*, pages 901–904, 2014.
- [45] C. Zitnick, Sing Bing Kang, Matt Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23:600–608, 08 2004.
- [46] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668, 2012.

- [47] Z. He, L. Yu, X. Zheng, S. Ma, and Y. He. Framework of avs2-video coding. In *2013 IEEE International Conference on Image Processing*, pages 1515–1519, 2013.
- [48] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje. The latest open-source video codec vp9 - an overview and preliminary results. In *2013 Picture Coding Symposium (PCS)*, pages 390–393, 2013.
- [49] J. Xu, R. Joshi, and R. A. Cohen. Overview of the emerging hevc screen content coding extension. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(1):50–62, 2016.
- [50] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand. 3d high-efficiency video coding for multi-view video and depth data. *IEEE Transactions on Image Processing*, 22(9):3366–3378, 2013.
- [51] G. Tech, Y. Chen, K. Müller, J. Ohm, A. Vetro, and Y. Wang. Overview of the multiview and 3d extensions of high efficiency video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(1):35–49, 2016.